DOCUMENT RESUME

ED 222 900 CS 207 220

AUTHOR Smith, Eugene
TITLE Evaluation of Writing with Semantic Differential Scales.
PUB DATE Apr 82
NOTE 15p.; Paper presented at the Annual Meeting of the Northwest Regional Conference of the National Council of Teachers of English (Boise, ID, April 22-24, 1982).
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Criteria; *Evaluation Methods; Higher Education; Peer Evaluation; Revision (Written Composition); *Semantic Differential; *Writing Evaluation; Writing Instruction

ABSTRACT

The semantic differential scale developed by Charles Osgood, George Suci, and Percy Tannenbaum can be modified to create a method of evaluating student writing that is clear, efficient, flexible, and specific. Based upon the premise that most terms have several dimensions of meaning, the theory creates scales by drawing "lines" through the "semantic space" of a term. In a writing class, discussion, stimulated by teacher prompting and guidance, can be used to develop appropriate criteria and evaluation scales for judging assignments. For instance, "evident" and "obscure" might serve as the evaluation scales for two criteria of "purpose." After receiving copies of the evaluation scales they have evolved, students engage in peer review. The teacher then checks off his or her own evaluation according to the scales and comments on the major strengths and weaknesses of the paper and its writing technique. Students may always use the evaluation to guide revision of any paper. All in all, this system has five advantages over conventional forms of evaluation: (1) the grading process is no longer mysterious, (2) student responsibility is increased, (3) evaluation is integrated into the writing-rewriting process, (4) arbitrariness is decreased, and (5) flexibility is increased. (Includes a sample evaluation scale.) (JL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED222900

EVALUATION OF WRITING WITH SEMANTIC DIFFERENTIAL SCALES

Eugene Smith
University of Washington

If we believe that good composition instruction should assist students at all stages of the process, how can we make evaluation of the product a convincing part of the process? Can teachers continue to treat final drafts as they have typically done--making marginal corrections, writing a summary comment, assigning an overall grade, returning the essays to students--perhaps with some class discussion--and then moving on to the next writing assignment?

As I became increasingly uncomfortable with the accustomed procedure, I felt compelled both to reconsider the purposes of essay evaluation and to develop procedures that would be better integrated with strategies for guiding the whole writing process. That reconsideration led me to articulate for my own guidance a set of imperatives:

1. You must make clear to students how you arrived at your judgments of quality.

2. Your evaluation of each essay must convincingly combine judgments of qualities specific to that essay with "universal" qualities of good expository writing.

3. Your comments and corrections must be related in tone, number, and complexity to your sense of each writer's capacity to

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Eugene Smith
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS207220

understand and use those comments.

4. Your evaluation should be well-organized and easy for the student to follow.

5. You must present the evaluation information legibly and efficiently.

6. Your evaluative comments should lead the student to see how the essay could be revised, and they should be generalizable to the student's next writing assignment.

None of my own writing teachers had come close to meeting these standards. Most of the writing evaluations I have seen from other teachers flouted nearly every one of them. Some of the worst--because they are so bewildering to students--involved only the affixing of a grade (or two grades, one for "ideas" or "content" and one for "mechanics"), with perhaps a word or two of praise, condolence, or castigation. Others, presumably from conscientious but prolix instructors, filled nearly a page and served as both discussion of ideas in the essay and commentary on writing technique. And, probably most important, seldom were students directly included in the process of determining and stating the criteria the instructor used for evaluation.

The method I have devised owes something to other familiar evaluation scales. Paul Diederich's analytic scale, developed by factor analysis of criteria applied to writing by practicing writers, offers both categories of criteria and descriptions of degrees of success. It is better adapted to use by raters of large numbers of anonymous essays than to students in one's class, however, and is therefore a summative rather than a formative evaluation. Any pre-

determined listing of criteria for writing quality probably contains that inherent flaw. The statements seem depersonalized; the numbers which express the evaluator's judgment seem harsh and difficult to interpret in relation to a student's desire to understand and improve.

I wanted a scale that had the clarity and efficiency of the Diederich scale but one that was both more flexible and more specific. I found the germ for it in the semantic differential scale developed by Osgood et al. to measure meaning.

Charles George Percy
Osgood, Suci, and Tannenbaum, in their pursuit of "the imps that inhabit the nervous system," focused on one of the most elusive--meaning. (The Measurement of Meaning, University of Illinois Press, 1957, p. 10) They started with a procedure familiar to any teacher: asking their subjects what a word like sophisticated means. The more fluent subjects could express the dimensions of meaning with reasonable depth. "It's being clever and wise about people and things--knowing the ropes, so to speak. t's sort of smooth and polished, graceful but [sic[not awkward . . . poised, 'savvy,' you know . . ." (p. 18) Such linguistic output has high validity, the investigators presumed, at least as much as any other introspective technique for finding out what meanings someone knows.

But when subjects' fluency was weaker, this "encoding" or expression of meanings for a term or concept was less spontaneous and probably less representative of their actual grasp of meaning. A method was needed, therefore, to elicit a large and representative sample of meanings that tend to define each term. The method is based on the premise that most terms (exceptions probably being words like the, and, after--structure words as opposed to words with notional mean-

ings) have several dimensions of meaning. Many of those dimensions are shared by most speakers of a language, however, so that a sample of meaning categories can be devised that includes the major ways in which meanings vary.

This cluster of dimensions of meaning is called "semantic space." While it cannot be fully known or measured, it can be tapped by drawing a series of "lines" through it, as though it were Euclidean in character. These "lines" become scales whose ends are defined by polar (opposite in meaning) adjectives.

Each line is sub-divided into seven parts in order to allow a user of the scale to register both direction and intensity of judgments. The summation of these judgments is then taken to be the "meaning" of the concept for the respondent.

In their discussion of the research underlying this method of measuring meaning and its application in various situations, the authors point out that it is "a highly generalizable technique of measurement." (p. 76) Eschewing a claim for "standard concepts" or "standard scales," they assert that each research problem will determine the class and form of the concept. In supporting a claim that polar adjectives are the best definers of semantic space, moreover, they cite the work of ethnolinguists in believing that "semantic opposition is common to most, if not all, language systems." (p. 327) Adjectives, as the means for expressing judgments, "are merely the most general and natural qualifiers in English." (p. 328)

During the six years I have been using such scales for evaluation of writing, I have evolved both their form (e.g., reducing the scale from seven to five parts) and the procedure for involving students with

them. From the first uses, however, students have taken well to them, saying essentially that this method of evaluation, though they have never experienced its use before, gives them reliable and useful information. Complaints about grades have been non-existent because students think they can readily see the basis for their grades. When the scales are combined with a policy permitting and encouraging rewriting of essays and then reevaluation before the end of the quarter, the evaluation scheme has even greater impact on their desire to improve writing and their will to continue working on it.

Arriving at Criteria and Evaluation Scales

During the composition process for each assigned essay, I devote at least one full class session to group discussion of criteria and evaluation scales. Working at the chalkboard, I ask students to suggest orally some criteria that seem to them especially important to judging that assignment.

One of my recent assignments, for example, requires the writing of a report based principally on information derived from an interview. Students come to realize that we are striving for more than a mere transcript of the exchange. We seek a report which integrates the information from the interviewee with a purpose determined by the writer. Asking why a reader might want to read the information is one of the tactics for inducing one or more possible writer's purposes. Remembering that they are supposed to make their purposes prominent in the finished essay, students readily suggest "Purpose" as one of the main categories of criteria.

Then I ask, "What are some specific qualities or aspects of purpose that a reader may perceive or expect? How can we state them briefly? Let's try to begin the statements of criteria with a verb." A recent class agreed on these two criterion statements of "Purpose": 1) expresses some aspect of interviewee's information; 2) conveys a sense of the relevance of the information. As we discussed the first, we said that, given the original formulation of the assignment which required that an interview be the principal source of information, it seemed reasonable to expect the writer not to conceal that source or make it seem so insignificant that the reader failed to get a strong sense of the interviewee's contribution. In other words, the writer must not wholly crowd out his/her source.

Our discussion of the second criterion emphasized that the reader must clearly understand why the interviewee's information seemed significant; the writer must shape it and perhaps comment on it so that a reader did not feel impelled to say "So what?" after reading the report.

As we formulated each criterion, we determined the appropriate evaluation scale. I explained that I thought we would find specific adjectives more stimulating to our judgment process and therefore more useful to each writer than general words like "good," "strong," or "weak." If we searched for single-word adjectives or short adjective phrases, we could arrive at a clearer sense of our standards of success for each criterion. Furthermore, if we tried to find adjectives that accurately expressed the probably extremes of quality, we would adequately describe the foreseeable range of quality.

One group of students decided that "evident" and "obscure" would work as evaluation scales for both criteria of "Purpose" cited above.

The class proceeded to other criteria, determining that "Organization" and "Audience" were necessary criteria categories. They decided on this statement for the former: "Has transitions between beginning, middle, and end to achieve unity." The evaluation scale went from "coherent" to "incoherent" (with the acknowledgment that "incoherent" need not mean incomprehensible). They arrived at two statements of criteria for "Audience": "Adapts language choices for assumed reader" and "Uses topic suited to audience." Evaluation scale adjective pairs for each were "sincere" (a word choice that surprised me because students these days generally don't want to talk much about sincerity) and "patronizing," "directed" and "diffuse."

This first attempt at devising criteria and scales was purposely brief so as not to overwhelm students. Lists of criteria and matching evaluation scales lengthened later in the quarter as students gained more experience in using them. Starting from a set devised for the first assignment, we reexamined each criterion category and statement to see whether it was still truly applicable to the present essay assignment. A process essay, they determined, called for deletion of the first "Purpose" criterion (cited above) and the addition of a second "Organization" criterion, namely, "Moves from step to step clearly," with the scale going from "clear" to "confused."

When we got to an essay that involved formulation and support for an argument, the categories expanded to include "Evidence/Support" and the total number of criteria swelled to ten.

While I admit to some prompting and fairly direct guidance in the

formulation of clear and succinct statements, students did most of the intellectual work of ascertaining criteria and trying to catch the particular qualities that would best characterize attainment of each criterion. They found that this is not easy to do, that the best choices of adjectives are often not immediately obvious. But because the process of arriving at the statements was public--and arduous--they understood them, saw their pertinence, and apparently considered their later use to be fair.

Using the Criteria and Evaluation Scales

After we had concluded our evaluation criteria discussion, I copied what I had written on the board for later duplication. The resulting dittoed sheets (see sample at end of article) were available in class for the session designated as evaluation time for the final draft (or "next-to-final draft," as I usually called it to emphasize the probable need for further rewriting). Students had understood from the outset of the course that they would be sharing one another's writing at several stages of completion, including a formal evaluation.

Collecting all the papers (all drafts including the final in a manila folder supplied by each student), I then redistributed the papers along with a copy of the evaluation sheet and asked each student to perform an evaluation on a peer's essay. I asked them to place a check on the scale for each criterion after they had read the essay at least twice and then to support each judgment with explanation on the lower part and back of the sheet. I stated emphatically that I considered this evaluation process a valuable part of their learning about writing

for two reasons: 1) having to formulate a considered and reasoned judgment helps to make one a more careful, more constructively critical reader; 2) each student would get the benefit of careful response from two people, a fellow student and the instructor, and could compare those responses for validity and later usefulness in rewriting.

That part of the evaluation complete, I invited each writer to read fellow students' responses immediately, if each wished to do so. Most showed curiosity about the responses, but any reaction to them was expressed informally rather than made a subject for general class discussion.

My time for evaluation came next, outside of class. I first read each essay and looked over the preliminary drafts to see what strategies each writer was using. I wanted to see whether there was evidence of use of invention strategies we had discussed, whether there were several early drafts, whether extensive revision seemed to be going on. My observations on these matters become a part of my written comment on the essay. I was finally ready to place my checks on the evaluation scales--often disagreeing with the peer readers' judgments. It was at this point that my sense of what constitutes excellence, mediocrity, and failure for college writing impinged most prominently. I felt no need to be guided by a student's comparative judgment when I knew that my judgment had a basis more firmly and extensively attached to experience in evaluating writing. I have consistently found, though, that having to make each of these separate evaluations according to specific criteria forces me to be especially thoughtful and to ask myself, "Can I support this judgment?"

After placing my checks on the scale divisions, I began the most important part of the formulative evaluation: writing comments. Two kinds are needed: a summary of major strengths and weaknesses and detailed observations on writing technique. The former are expansions on and support for several of the judgments registered on the semantic differential scales. The latter offer such assistance as changes in diction, revisions in sentence structure, questions about adequacy and suitability of evidence, and corrections in spelling and punctuation. (A numbering scheme for these detailed comments works well, the number for each comment matching a number handwritten on the manuscript.)

An example of one student's handling of the report essay will show how the semantic differential scale judgments lead to comment. The student had interviewed a man who occasionally sold handmade key chains on the university campus. The man had had a colorful life as a seaman and apparently held strong opinions on many subjects. The essay tried to catch the flavor of his talk and to condense several of his anecdotes, but its only principle of organization was a sketchy narrative frame and a series of anecdotes and descriptive details. It also lacked a strong sense of purpose. A student evaluator rated the essay at the mid-point on the scales for organization, purpose and clarity, and coherence. My ratings on those criteria were nearly the same.

The student's comments included the following: "I had the sense of an organizing principle present in the paper, but I had no point of entry . . you may assume too much knowledge (or too much insight) on the part of your readers. Perhaps if you looked at this paper

from the perspective of someone coming to it 'cold' (as in a newspaper), you could see some of the problems I had in trying to knit Frank's dialogue, the "preachers," and the educational and financial establishment into a cogent account."

My summary comments included these observations: "Both by the selection of a title ['Yarn Spinner'] and the recounting of anecdotes, you manage deftly to give a character sketch of Frank. You present him as a surprisingly complex man who can entertain and perhaps instruct. What I most miss, though, is a point of view from you. Where do you stand in relation to him? If your intention was merely to report a few of his stories, can you suggest overtly to the reader why those stories may be worth reading? If you have some observation to make about this yarn spinner--such as what we might learn from him--that might be an even better thread to run through the whole article, giving it a stronger sense of purpose and orienting it more clearly toward a certain audience."

These observations were futher specified in my numbered comments, such as "This paragraph seems to consist of loosely related details. Do they add up to some unified idea or impression?" "Here I feel that you are yanking me to a different subject without showing me why." "Do you understand what this statement means? I don't." "Here is one of the places where you seem to be making a judgment about him. It is that sense of judgment--elaborated--that I think you might use to give organizational fiber to this article."

The writer therefore got a substantial amount of evaluative information: ratings by clearly established criteria, judgments about major strengths and weaknesses, questions and suggestions about

specific strategies, and guidance for revision.

The final evaluative act may be to affix a grade. Since the five-part scale naturally lends itself to interpretation on an A to F continuum, it would be possible simply to base the grade on an average of these scale evaluations. But since none of our class-derived, dittoed criteria/scale arrays pretends to be complete, I feel no inherent limitation to base my overall evaluation on those criteria. By the time students read the grade, however, they have seen so much detailed reasoning and evaluation that went into it that they tend to be convinced of its validity.

Relation of Evaluation to Revision

Since I tell students during the quarter that no piece of writing is ever really finished--in that it can always be improved, even by the best professional writer--they know that my detailed evaluation can be used as a guide to revision. I also tell them, as encouragement to use it that way, that I will reread and reevaluate any of the essays they wish to revise and that the final grade for the course will reflect the further improvement that revised essays demonstrate. Some students undertake extensive revision of at least one essay, depending primarily on available time near the always hectic end of the quarter.

Summary of Major Advantages

Using a semantic differential scale system for evaluating students' writing has these major advantages over the more conventional forms of

evaluation:

1. Much of the mystery of how a teacher arrives at judgments about the quality of writing tends to evaporate. (That may be a slight disadvantage for the teacher who wishes to play sphinx, inscrutable keeper of the riddle of the grade.)

2. Students assume more individual responsibility for thinking about and articulating the qualities that define good writing, internalizing them as all effective, independent writers must do.

3. Evaluation becomes an integral part of the writing-rewriting process, rather than a climax without effectual follow-up.

4. The criteria and means for applying them achieve a degree of explicitness and fairness that takes much of the arbitrariness out of evaluation.

5. The method is highly flexible, allowing adaptation for an enormous variety of kinds of writing assignments and for varying levels of students' experience and sophistication with writing.

EVALUATION OF ESSAY: ARGUMENT IN SUPPORT OF OPINION

Writer_____ Evaluator_____

Criteria	Evaluation Scales
Organization:	
Leads reader coherently through logic which supports the position	coherent ——— disjointed
Has plausible beginning, middle, and end	distinct ——— indistinct
Integrates generalizations and evidence	integrated ——— choppy
Purpose:	
Persuades reader that opinion is valid or justifiable	convincing ——— implausible
Leads persuasively to a focused conclusion	focused ——— hazy
Audience:	
Seems pertinent to present-day readers	clearly pertinent ——— doubtfully pertinent
Adopts a tone suited to topic and audience	fitting ——— inappropriate
Evidence/support:	
Presents evidence in clear, easily understood, persuasive manner	sharp ——— vague
Uses evidence/support suited to topic	suitable ——— unsuitable
Uses enough evidence/support	sufficient ——— insufficient